

1/26

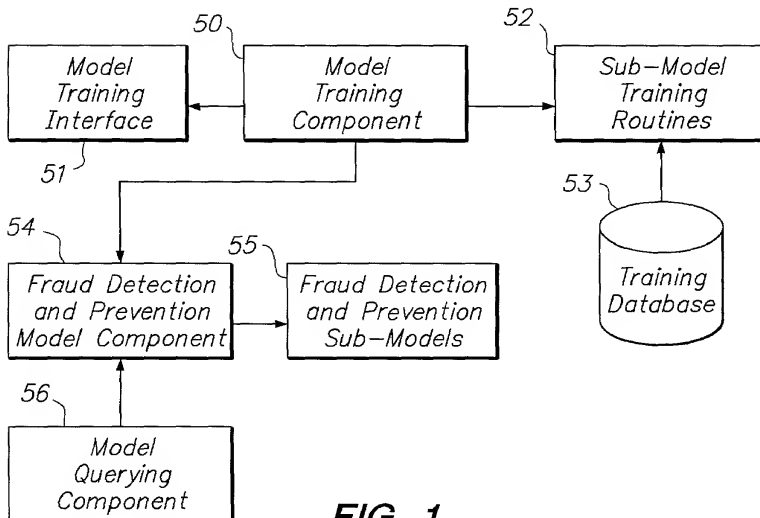


FIG. 1

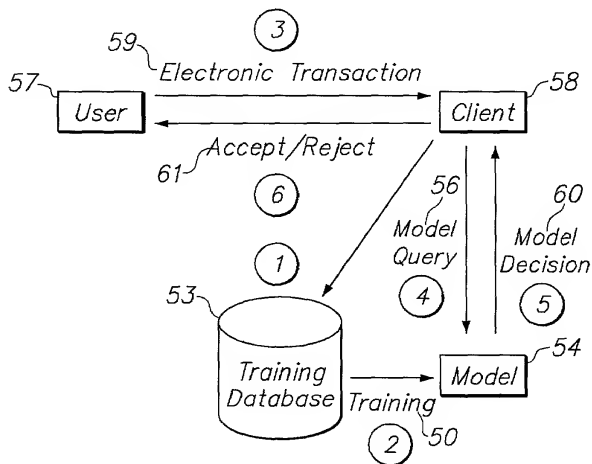


FIG. 2

09810313.073001

2/26

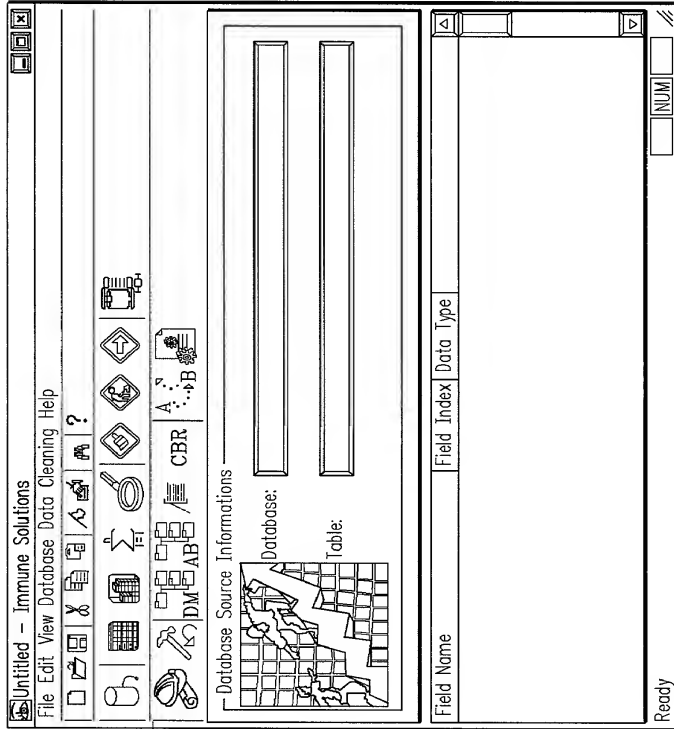


FIG. 3

3/26

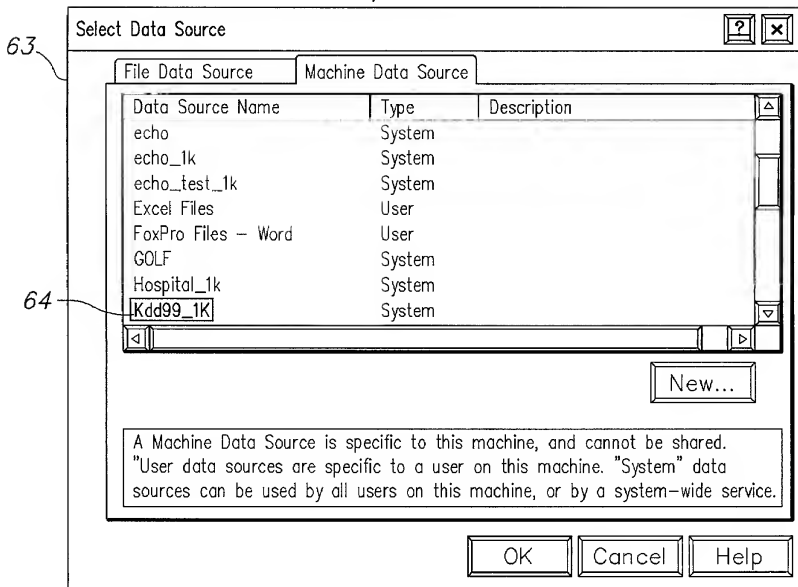


FIG. 4

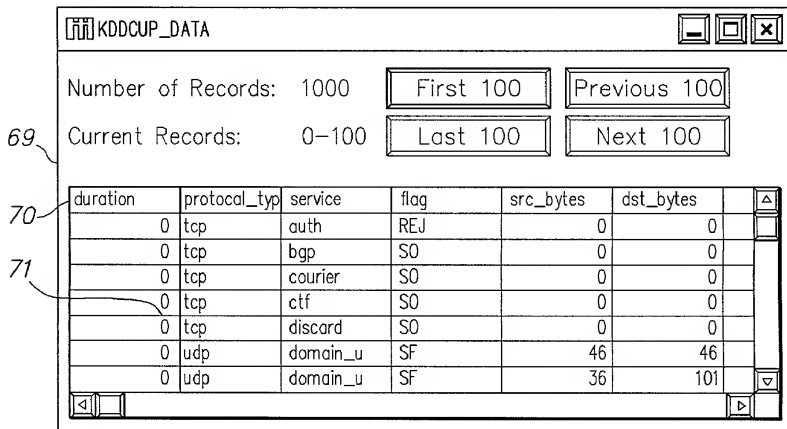


FIG. 6

09810313 073001

4/26

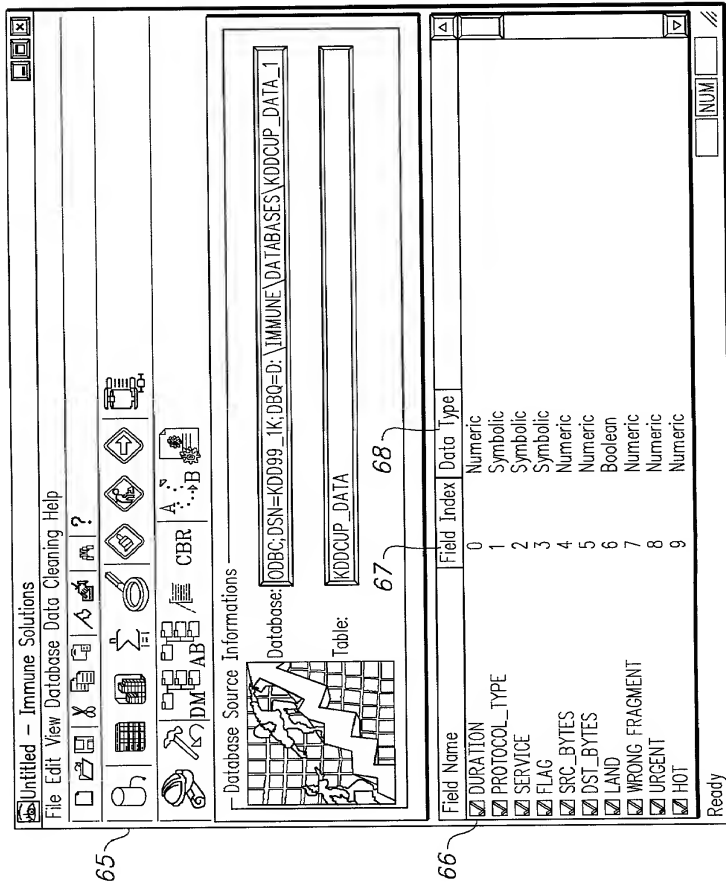


FIG. 5

5/26

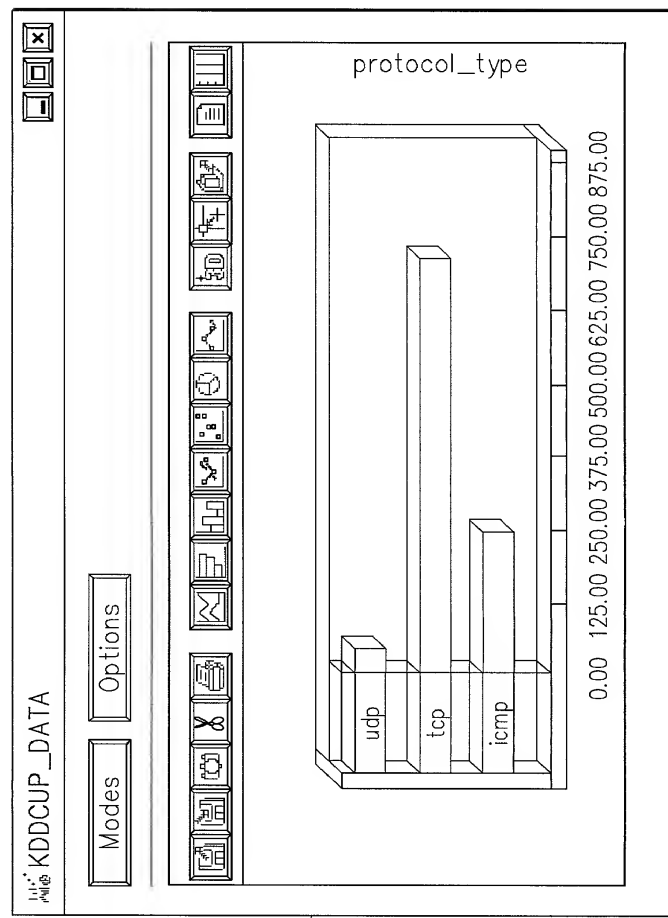


FIG. 7A

100E20" ETEOT860

KDDCUP_DATA

Modes

Options

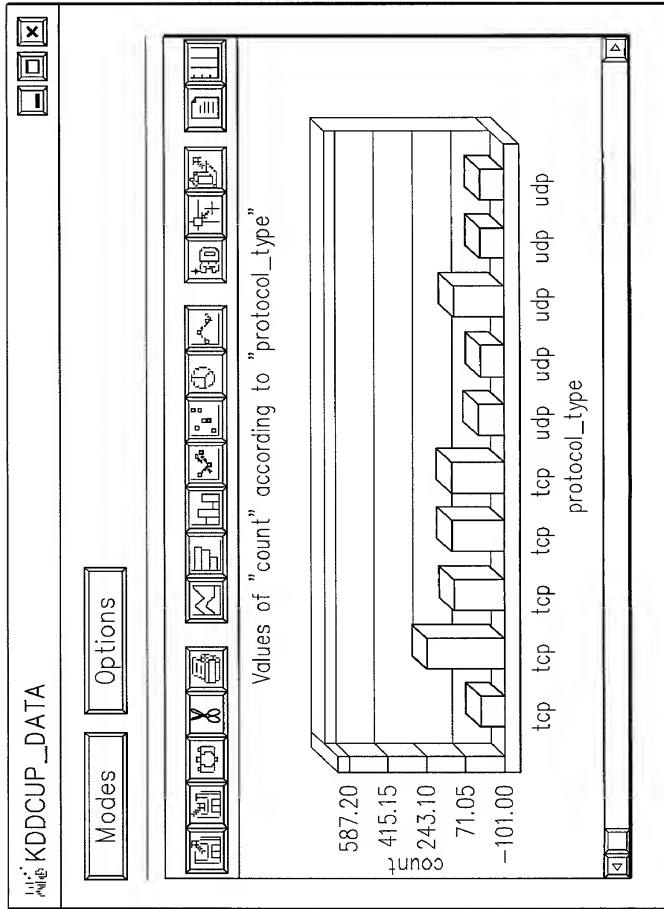


FIG. 7B

7/26

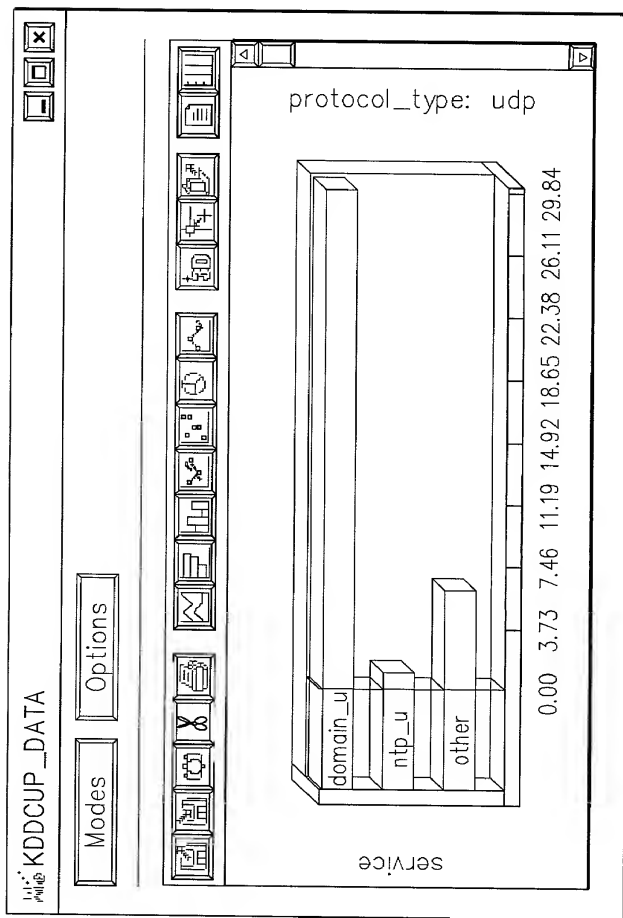


FIG. 7C

8/26

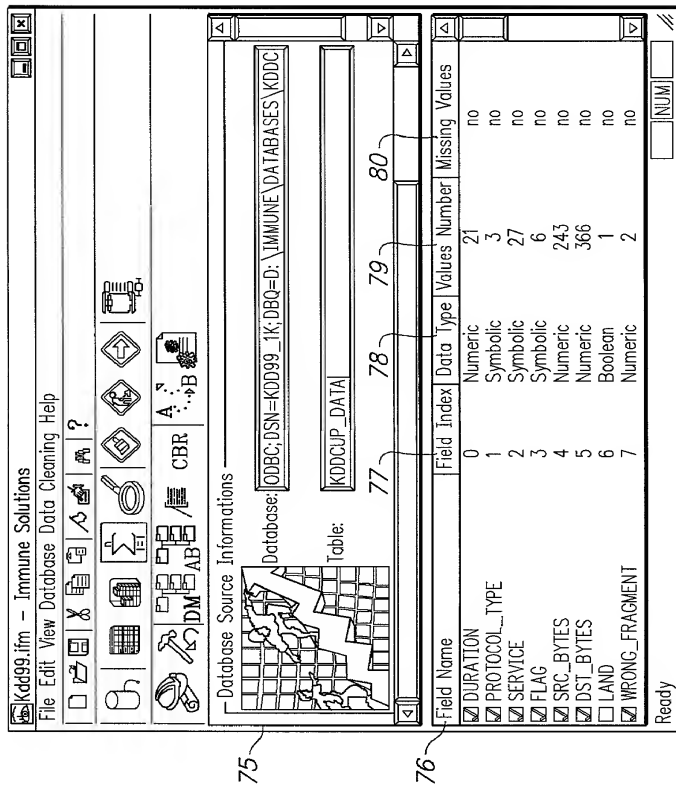


FIG. 8

9/26

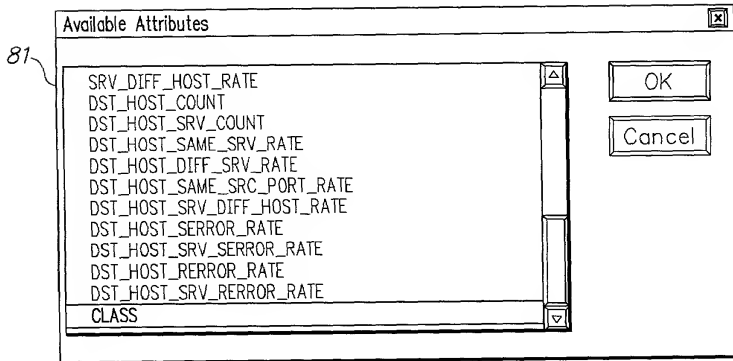


FIG. 9

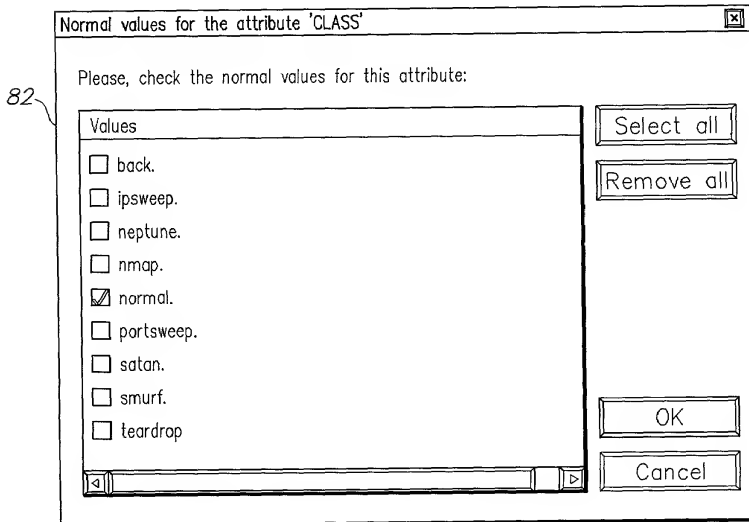


FIG. 10

10/26

FIG. 11

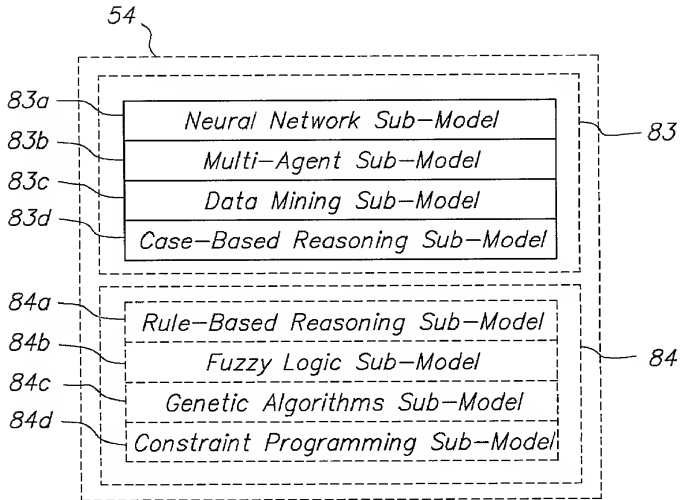
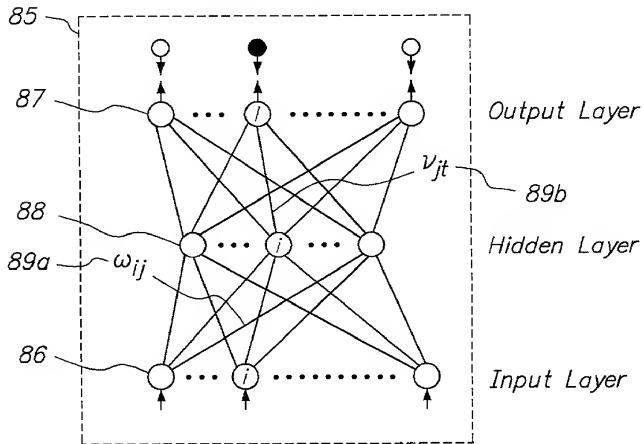


FIG. 12



09810313.073001

11/26

FIG. 13

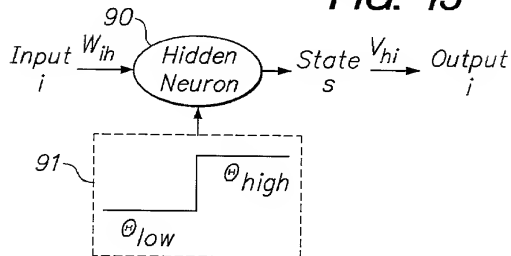
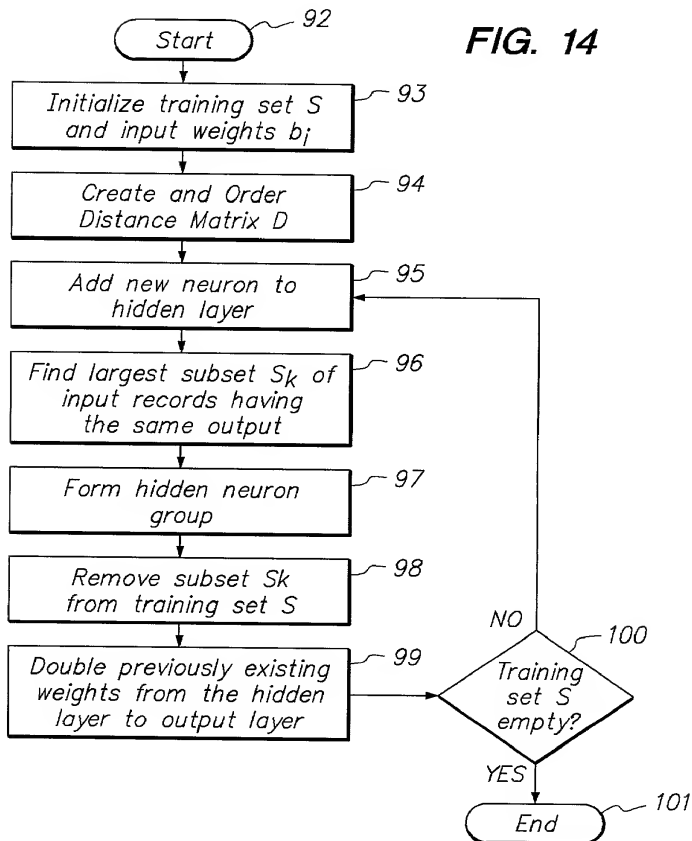


FIG. 14



09810313.073001

12/26

102a	<i>Euclidean</i>	$d(X_i, X_j) = \sqrt{\sum_{k=1}^M (X_{i,k} - X_{j,k})^2}$
102b	<i>Manhattan</i>	$d(X_i, X_j) = \sum_{k=1}^M X_{i,k} - X_{j,k} $
102c	<i>Normalized Euclidean</i>	$d(X_i, X_j) = \sqrt{\frac{1}{M} \sum_{k=1}^M \left(\frac{X_{i,k} - X_{j,k}}{\max_k - \min_k} \right)^2}$
102d	<i>Normalized Manhattan</i>	$d(X_i, X_j) = \frac{1}{M} \sum_{k=1}^M \left \frac{X_{i,k} - X_{j,k}}{\max_k - \min_k} \right $
102e	<i>Weighted-Euclidean</i>	$d(X_i, X_j) = \sqrt{\sum_{k=1}^M b_i * (X_{i,k} - X_{j,k})^2}$

FIG. 15

09810313 073001

13/26

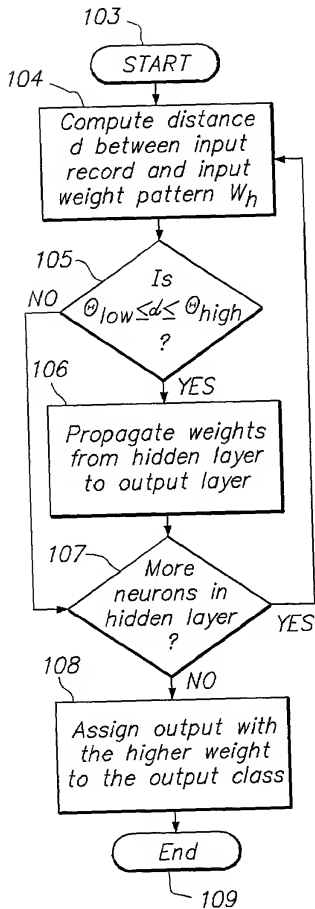


FIG. 16

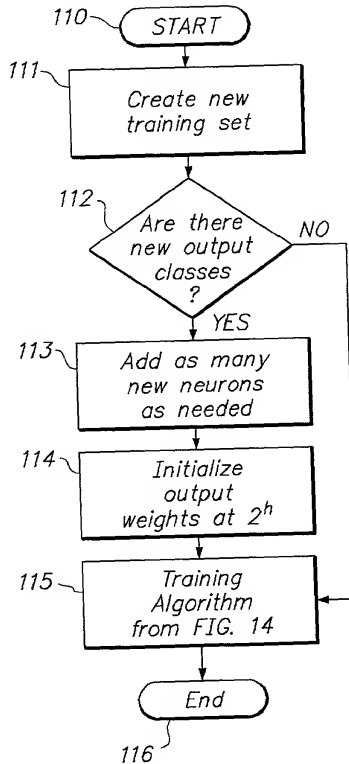
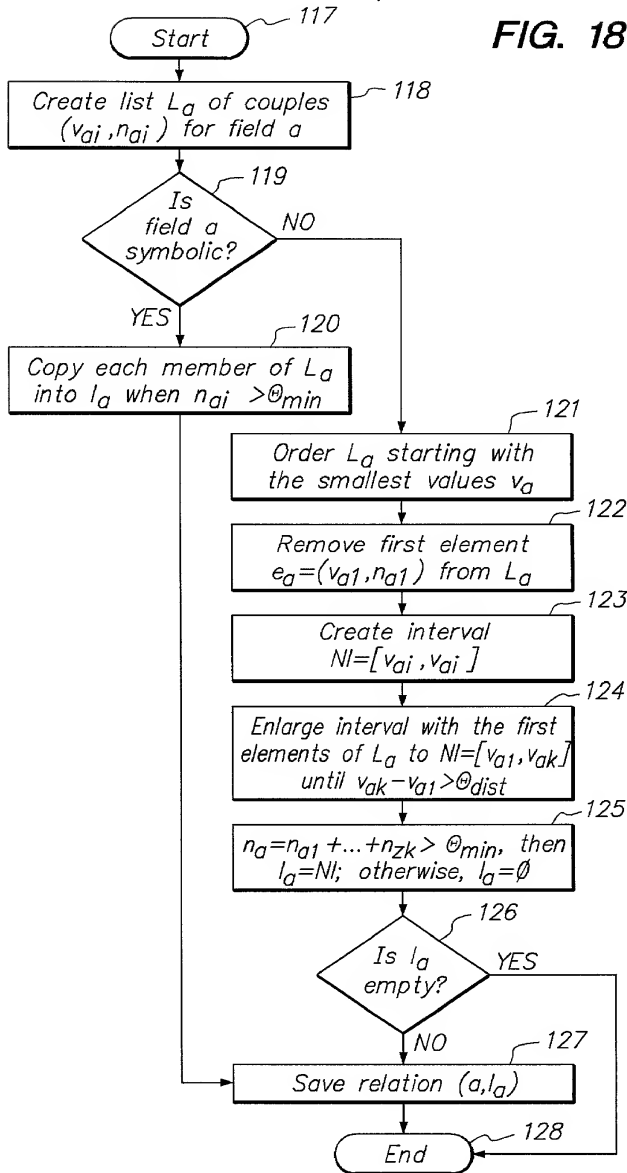


FIG. 17

14/26

FIG. 18



09810313.073001

15/26

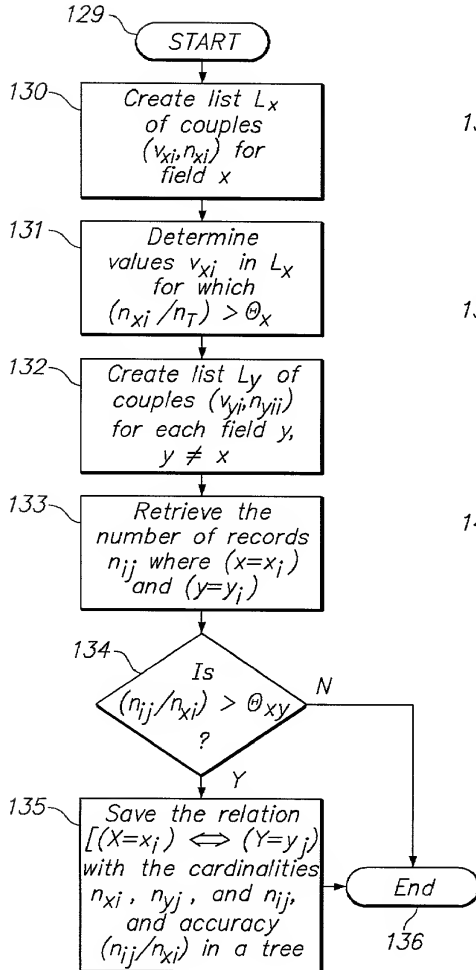


FIG. 19

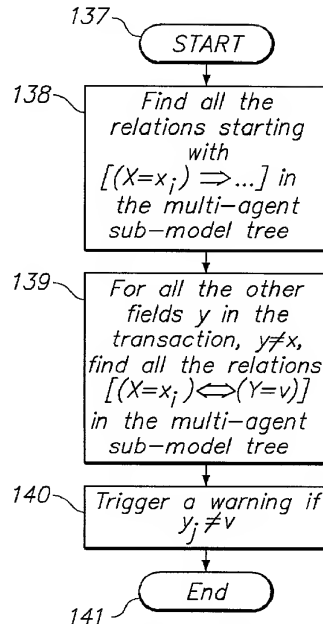
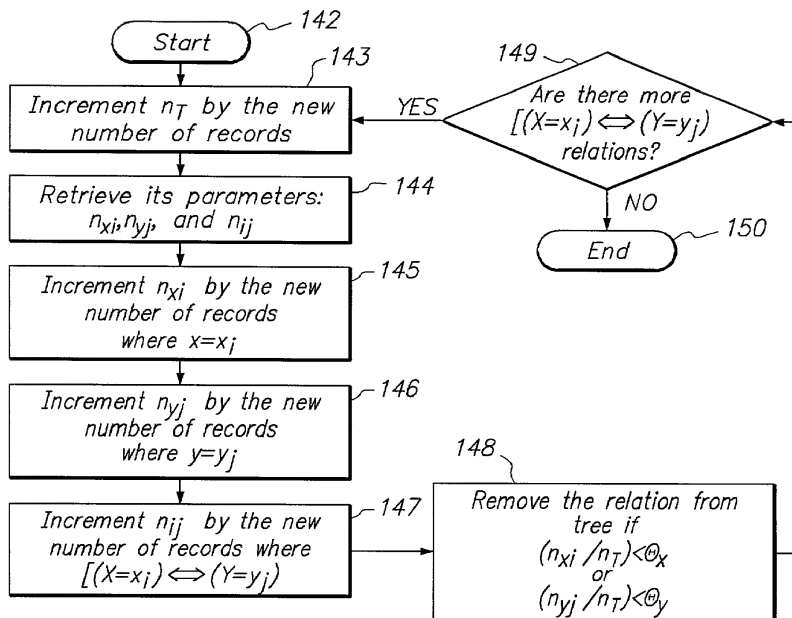


FIG. 20

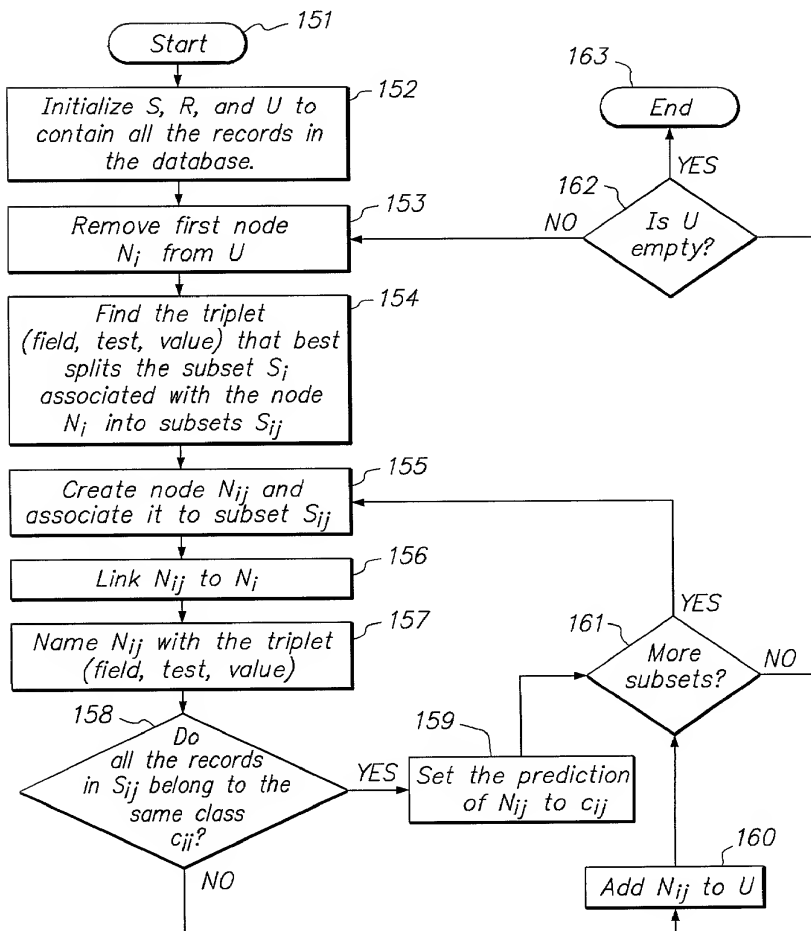
09/810,313 07/3001

16/26

FIG. 21



09810313 073001



18/26

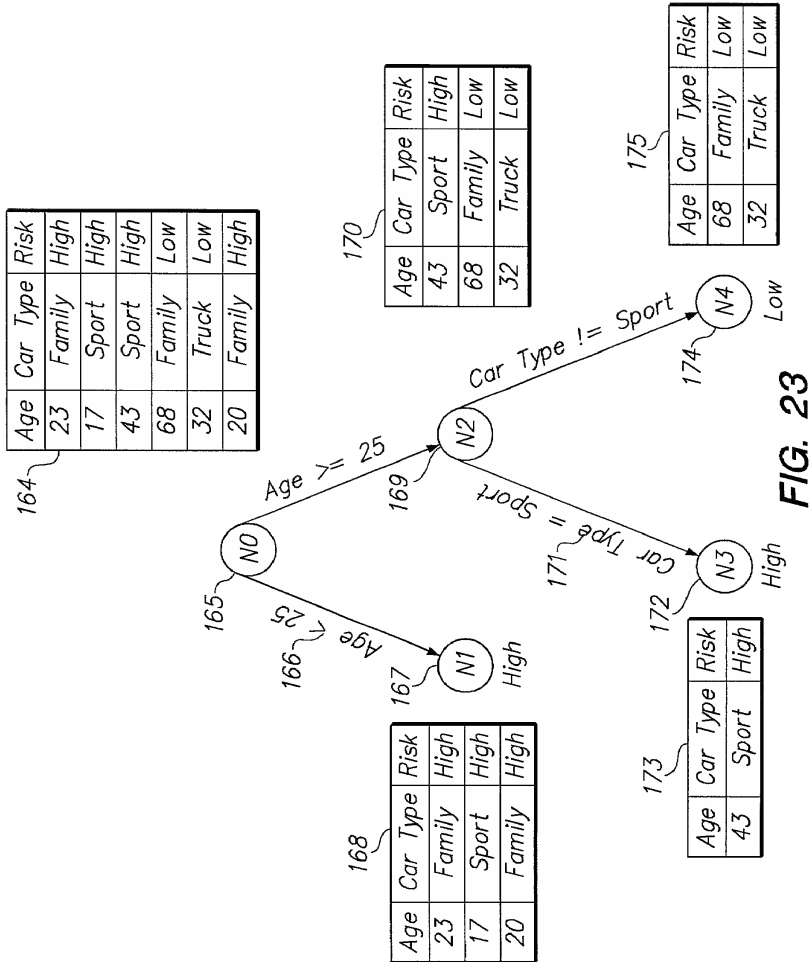


FIG. 23

19/26

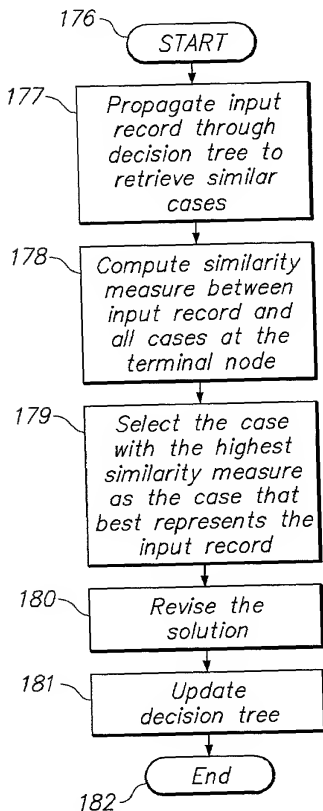


FIG. 24

20/26

183

<i>Global Similarity Measure</i>	<i>Expression</i>
<i>City-block</i>	$\frac{1}{\rho} \sum_{i=4}^{\rho} sim_i (V_{1i}, V_{2i})$
<i>Weighted city-block</i>	$\frac{1}{\rho} \sum_{i=4}^{\rho} w_i * sim_i (V_{1i}, V_{2i})$
<i>Euclidean</i>	$\frac{1}{\rho} \sqrt{\sum_{i=4}^{\rho} sim_i (V_{1i}, V_{2i})^2}$
<i>Minkowski</i>	$\frac{1}{\rho} r \sqrt[r]{\sum_{i=1}^{\rho} sim_i (V_{1i}, V_{2i})^r}$
<i>Weighted Minkowski</i>	$r \sqrt[r]{\sum_{i=1}^{\rho} w_i * sim_i (V_{1i}, V_{2i})^r}$
<i>Weighted maximum</i>	$max_i w_i * sim_i (V_{1i}, V_{2i})$

FIG. 25

184 21/26

Local Similarity Measures	Field Type	Field Valuation
$\begin{cases} 0, & \text{if } V_1 \cap V_2 = \emptyset \\ 1, & \text{otherwise} \end{cases}$	Nominal	Single, multiple
$\frac{\text{Card}(V_1 \cup V_2) - \text{Card}(V_1 \cap V_2)}{\text{Card}(V_1 \cup V_2)}$	Nominal	Multiple
$\frac{\text{Card}(V_1 \cup V_2) - \text{Card}(V_1 \cap V_2)}{\text{Min}(V_1 \cup V_2)}$	Nominal	Multiple
$\frac{\text{Card}(V_1 \cup V_2) - \text{Card}(V_1 \cap V_2)}{\text{Max}(V_1 \cup V_2)}$	Nominal	Multiple
$\frac{\text{Card}(V_1 \cup V_2) - \text{Card}(V_1 \cap V_2)}{\text{Card}(0)}$	Nominal	Multiple
$\frac{\text{ec}(\text{min}(V_1^-, V_2^-), \text{max}(V_1^+, V_2^+)) - \text{Card}(V_1 \cap V_2)}{\text{Card}(0)}$	Ordinal, Numeric	Multiple
$\frac{ V_1 - V_2 }{\text{ec}(0)}$	Numeric	Single
$\frac{ V_{1c} - V_{2c} }{\text{ec}(0)}$	Numeric	Multiple
$\frac{\text{ec}(\text{min}(V_1^-, V_2^-), \text{max}(V_1^+, V_2^+)) - \text{ec}(V_1 \cap V_2)}{\text{ec}(0)}$	Numeric	Multiple
$\frac{\text{ec}(V_1 \cup V_2) - \text{ec}(V_1 \cap V_2)}{\text{ec}(V_1 \cup V_2)}$	Numeric	Multiple
$\frac{\text{ec}(V_1 \cup V_2) - \text{ec}(V_1 \cap V_2)}{\text{min}(\text{ec}V_1, \text{ec}V_2)}$	Numeric	Multiple
$\frac{\text{ec}(V_1 \cup V_2) - \text{ec}(V_1 \cap V_2)}{\text{max}(\text{ec}V_1, \text{ec}V_2)}$	Numeric	Multiple
$\frac{2 * h(V_1 \cup V_2) - h(V_1) - h(V_2)}{2 * h_{\text{max}}}$	Taxonomic	Multiple
$\frac{h(\text{node that unit } V_1 \text{ \& } V_2)}{\text{total height of } h}$	Taxonomic	Single

FIG. 26

09810313 073001

22/26

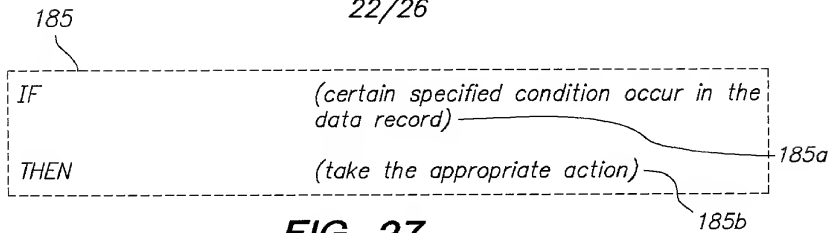


FIG. 27

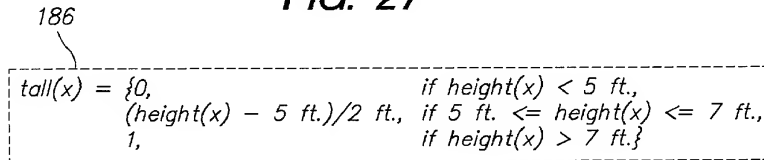


FIG. 28

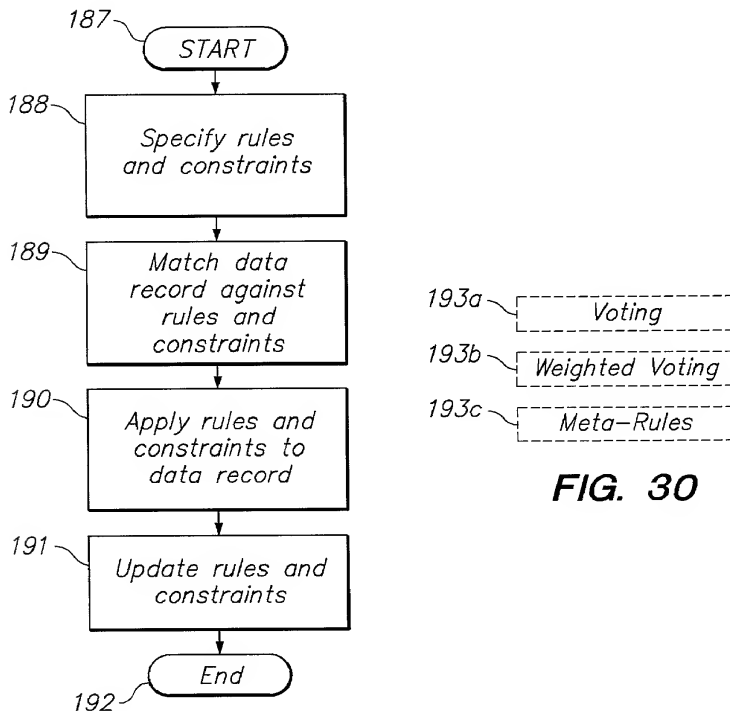


FIG. 30

FIG. 29

09810313 073001

23/26

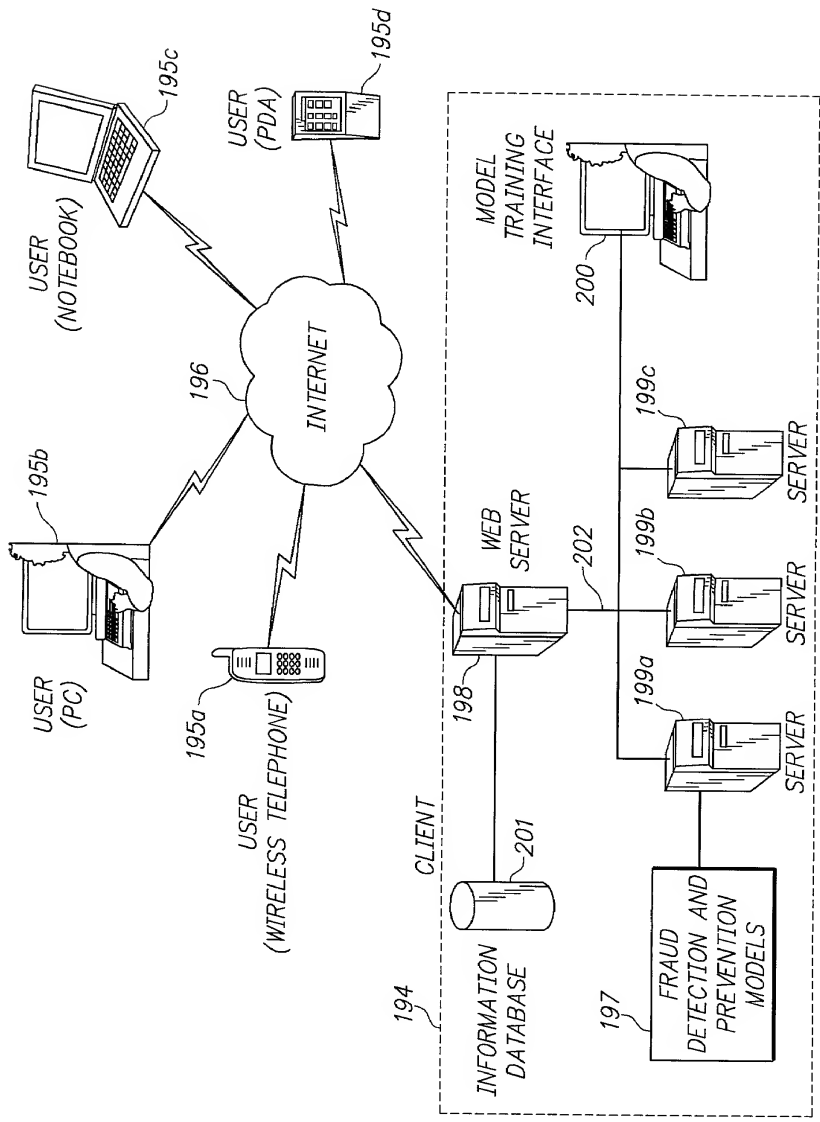


FIG. 31

FIG. 20

24/26

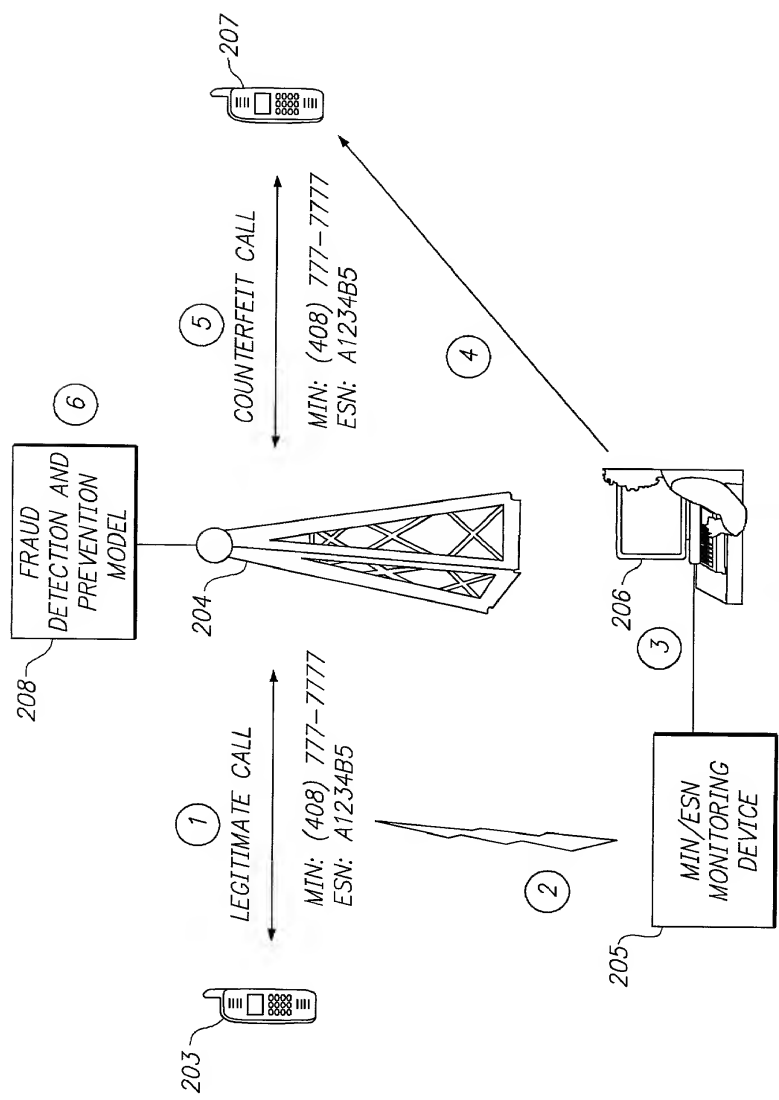


FIG. 32

09810313 00000000

25/26

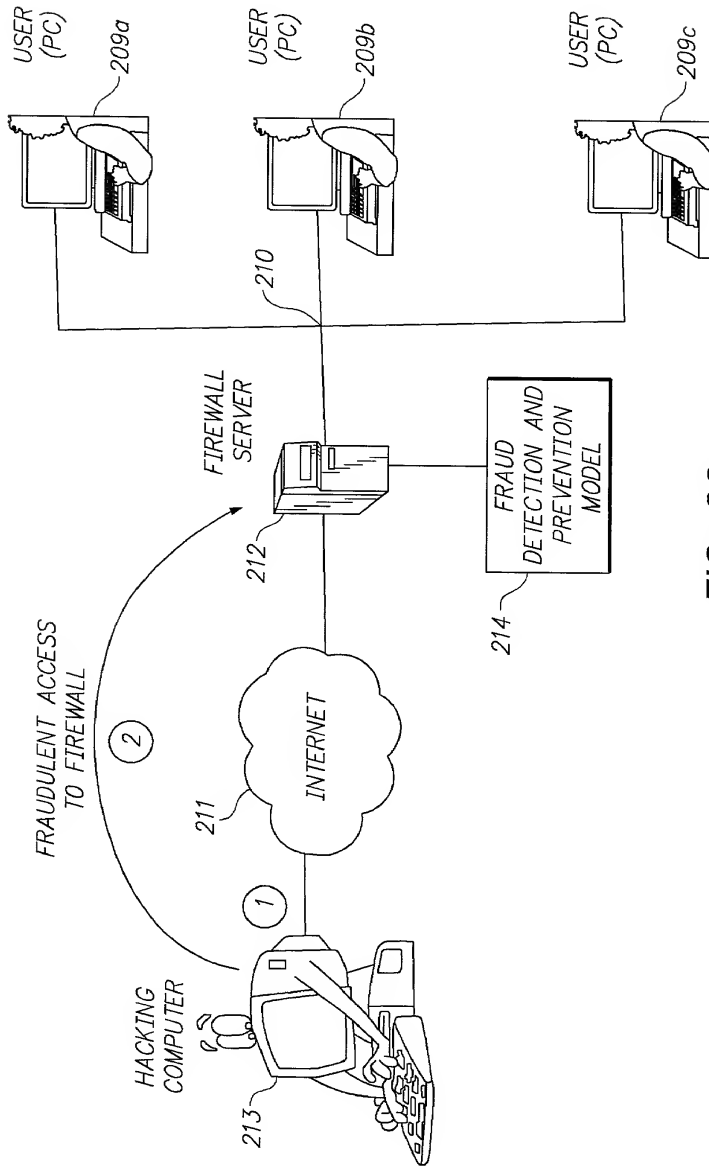


FIG. 33

215

26/26

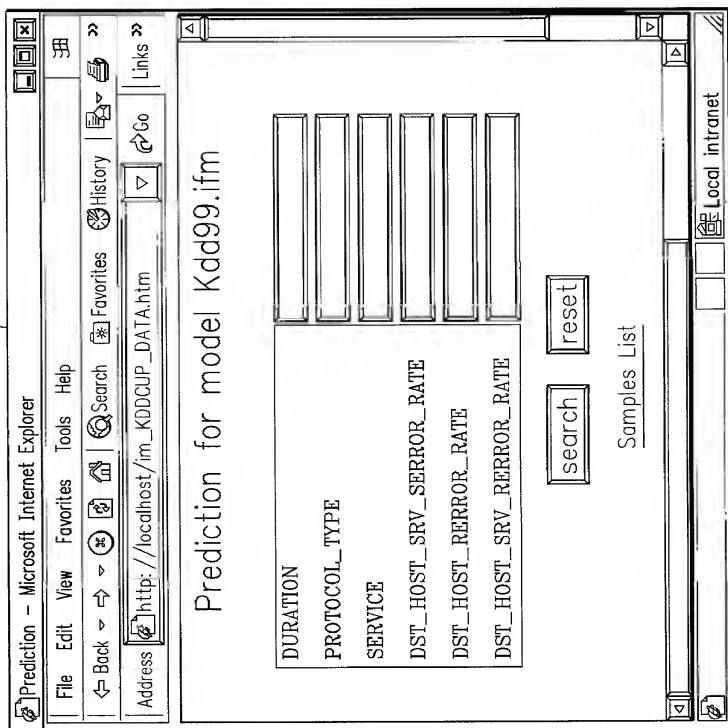


FIG. 34